Simple Regression Model

January 24, 2011

Outline

Descriptive Analysis

Causal Estimation

Forecasting

Regression Model

We are actually going to derive the linear regression model in 3 very different ways

While the math for doing it is identical, conceptually they are very different ideas

Outline

Descriptive Analysis

Here our goal is simply to estimate $E(y \mid x)$

However, when x is continuous we can't do it directly

thus, we need a model for what this looks like

The easiest model I can think of is

$$E(y \mid x) = \beta_0 + \beta_1 x$$

To interpret it notice that

$$E(y \mid x = 0) = \beta_0$$

you can also see that β_1 is the slope coefficient

$$\frac{\partial E(y \mid x = 0)}{\partial x} = \beta_1$$

Estimation

- Now we need some way to estimate it
- It will be useful to define

$$u = y - \beta_0 - \beta_1 x$$

- What do we know about u?
- The fact that $E(y \mid x) = \beta_0 + \beta_1 x$ means that

$$\begin{aligned} E(u \mid x) &= E(y - \beta_0 - \beta_1 x \mid x) \\ &= E(y \mid x) - \beta_0 - \beta_1 x \\ &= \beta_0 + \beta_1 x - \beta_0 - \beta_1 x \\ &= 0 \end{aligned}$$

The fact that $E(u \mid x) = 0$ means two important things:

1. $$E(u) = 0$$

2. The expected value of u does not change when we change x.
   This second thing has a number of implications, but the most useful (and intuitive) is that

$$\begin{aligned} 0 &= cov(u, x) \\ &= E(ux) - E(u)E(x) \end{aligned}$$

Putting this together we have two conditions:

$$\begin{aligned} E(u) &= 0 \\ E(ux) &= 0 \end{aligned}$$

What was the point of all of that?

To estimate an expectation we use a sample mean.

What if replace the expectations above with sample mean expressions?

Before doing this lets be clear what we mean the sample.

Assume that I observe a sample size of N

Let $i = 1, ..., N$ index the people in the data

Lets look at this for the voting data set

We will write the population regression model as

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

where (for example) y_i is the value of y for individual i.

Now lets go back to the equations.

To think about estimation lets define the sample regression model

$$y_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i + \hat{u}_i$$

where

- $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are our estimates of β_0 and β_1 from the sample
- $\hat{u}_i$ is defined as

$$\hat{u}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i$$

How do we want to estimate this model?

Well if $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are like β_0 and β_1, then $\hat{u}_i$ should be like u_i.

We know that in the population $E(u_i) = 0$ so it makes sense to force this to be approximately true in the sample

This is the idea of a sample analogue

If the sample looks like the population, then lets force things to be true in the sample which we know would be true in the population

$$
\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N}\widehat{u}_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i - \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_0 - \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_1 x_i \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i - \widehat{\beta}_0 - \widehat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i \\
&= \overline{y} - \widehat{\beta}_0 - \widehat{\beta}_1 \overline{x}
\end{aligned}
$$

Which I can write as

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

Now we have two unknowns an one equation, we need one more.

The population expression is

$$E(u_i x_i) = 0.$$

The sample analogue of this is

$$
\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N}\widehat{u}_i x_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) x_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[y_i x_i - \widehat{\beta}_0 x_i - \widehat{\beta}_1 x_i^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i x_i - \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_0 x_i - \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_1 x_i^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i x_i - \widehat{\beta}_0\frac{1}{N}\sum_{i=1}^{N} x_i - \widehat{\beta}_1\frac{1}{N}\sum_{i=1}^{N} x_i^2
\end{aligned}
$$

now lets take this equation and plug in the fact that $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$

$$
\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N} y_i x_i - \widehat{\beta}_0 \frac{1}{N}\sum_{i=1}^{N} x_i - \widehat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i x_i - \left(\overline{y} - \widehat{\beta}_1 \overline{x}\right) \frac{1}{N}\sum_{i=1}^{N} x_i - \widehat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i x_i - \frac{1}{N}\sum_{i=1}^{N} \overline{y} x_i + \widehat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i \overline{x} - \widehat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i x_i \\
&= \frac{1}{N}\sum_{i=1}^{N} \left(y_i - \overline{y}\right) x_i - \widehat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i \left(x_i - \overline{x}\right)
\end{aligned}
$$

Next we will use a result about means that we discussed in the statistics review

$$\begin{aligned}
\sum_{i=1}^{N} x_i \left(x_i - \overline{x}\right) &= \sum_{i=1}^{N} x_i \left(x_i - \overline{x}\right) - \sum_{i=1}^{N} \overline{x} \left(x_i - \overline{x}\right) \\
&= \sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2 \\
\sum_{i=1}^{N} \left(y_i - \overline{y}\right) x_i &= \sum_{i=1}^{N} \left(y_i - \overline{y}\right) x_i - \sum_{i=1}^{N} \left(y_i - \overline{y}\right) \overline{x} \\
&= \sum_{i=1}^{N} \left(y_i - \overline{y}\right) \left(x_i - \bar{x}\right)
\end{aligned}$$

Using that we can solve for $\widehat{\beta}_1$

$$\widehat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})(x_i - \bar{x})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

and we still know that

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

And we are done.

Note that since

$$\widehat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})(x_i - \bar{x})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

it is essentially the sample covariance between x and y divided by the sample variance of x.

The sample variance must be positive therefore:

- The regression coefficient, the covariance, and the correlation coefficient must all have the same sign
- They will only be zero if all of them are zero

Voting Example

Lets use this to think about the voting example

We can say nothing about causality.

There is a clear positive relationship, but it could be due to anything

$x \rightarrow y$	Spending a lot of money gets people to like you which leads to more votes
$z \rightarrow x, z \rightarrow y$	People who are popular attract a lot of money and get a lot of votes
$y \rightarrow x$	Its all about influence. I only give money to the people who are going to win so I can influence their future votes.

CEO example

As another example lets consider the relationship between CEO salary and the Return on Equity

This comes from the data set CEOSAL1 in Wooldridge

We know that CEOs make tons of money, but is it just a scam or are they actually worth something?

The key variable is return on equity (roe) defined as “net income as a percentage of common equity”

A number of 10 means if I invest a $100 in equity in a firm I earn $10 each year

Salary is measured in terms of $1000 per year

We will write the conditional expectation as

$$E(salary \mid roe) = \beta_0 + \beta_1 roe$$

We can run this regression in stata

For every point higher in return on equity CEO salary raised by \$18,501 per year

Really not that much if you think about it.

Is this causal?

Again we have all of the three possibilities

$x \to y$	When stock price does well, CEOs get a raise
$z \to x, z \to y$	Good CEOs tend to make a lot of money and their companies perform well
$y \to x$	By paying my CEO more I get her to work harder and push the stock price up

Outline

To really say something about causality we need to make some more assumptions

This involves writing down a “structural data generating model”

This will look similar to what we have been doing, but conceptually is very different

- For the conditional expectation data we started with the data and then asked which model could help summarize it
- For the structural case we start with the model and then use it to say what the data will look like

OK we need a model to do this

We will use the same regression model:

$$y = \beta_0 + \beta_1 x + u$$

and think about it this way:

1. β_0 and β_1 are real number which are out there in nature and we want to uncover them
2. you choose x
3. Nature (or God) chose u in a way that is unrelated to your choice of x

This is now a real causal model

suppose that

- y is your annual income
- x is your education
- $\beta_0 = 10,000$
- $\beta_1 = 5000$
- $u = -5000$

So what does this mean?

Suppose I choose different levels of education, here is the earnings I will get

Education	Earnings
0	$5000
12	$65,000
16	$85,000
18	$95,000

It might seem a little goofy to assume that *u* were known as you might think that you make your college going decision without knowing *u*

Actually, it kind of doesn't really matter

If you know the model you know that graduating college (16) versus not going to college (12) is worth about $20,000 per year

Probably a pretty good investment

What is u?

It is things that affect your earnings other than education.

Examples:

- Intelligence
- Work Effort
- Smoothness
- Family Connections
- race
- gender
- age

Estimation

Now we need a way to estimate our model In particular we want to estimate β_0 and β_1

The question is how is u determined?

The standard assumption (at least the one to start with) is that u is essentially assigned at random

We can write this as

$$E(u \mid x) = 0$$

Note that this is the same as what we did before, but conceptually very different

- Before u was defined simply as $y - \beta_0 - \beta_1 x$ it didn't actually mean anything
- Now we think of u as this real thing that is actually out there and means something-it is just that we can't observe it

As before, the fact that we have assumed $E(u \mid x) = 0$ really means two separate things one of which is a big deal the other isn't:

1. $E(u \mid x)$ does not vary with x
2. $E(u) = 0$ (as opposed to some other number)

The first is a really important assumption.

The second really isn't. Suppose we instead assumed that

$$E(u \mid x) = 6$$

with the model

$$\beta_0 + \beta_1 x + u$$

This is equivalent to another model

$$\gamma_0 + \gamma_1 x + \nu$$

with

$$\begin{aligned} \gamma_0 &= \beta_0 + 6 \\ \gamma_1 &= \beta_1 \\ \nu &= u - 6 \end{aligned}$$

Notice that now

- $$\begin{aligned} E(\nu \mid x) &= E(u-6 \mid x) \\ &= E(u \mid x) - 6 \\ &= 6 - 6 \\ &= 0 \end{aligned}$$

- β_1 and γ_1 are the same and that is really what we care about (remember the education model above)
- Thus, the fact that we pick zero is really a Normalization. It makes the model well defined (identified) without any real imposition on the data
- However, the fact that $E(u \mid x)$ does not vary with x is much more than a normalization. It has real content.

Now that we have a model how do we estimate it?

Now nothing is really any different than before

We can write down the sample regression function as

$$y_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i + \hat{u}_i$$

we know that

$$\begin{aligned} E(u_i) &= 0 \\ E(u_i x_i) &= 0 \end{aligned}$$

So it makes sense to use the sample analogues

$$\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i = 0$$

$$\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i x_i = 0$$

which will give you

$$\widehat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})(x_i - \bar{x})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

exactly as before

It is only the interpretation that has changed.

Lets look back on some of our models and interpret things in this way.

Let me be very clear about something: the question of whether this additional assumption is reasonable or not is definitely questionable in all of these cases

Lets worry about that later.

For now, just take the assumption as given and worry about the interpretation.

Voting Example

For the voting example we get the sample regression function

$$voteA = 26.812 + 0.463 shareA + \widehat{u}_i$$

As usual the $\widehat{\beta_0}$ parameter is not very interesting. It tells me the fraction of the votes I would get if I spent no money.

I would get about 27% of the vote on average if I spent no money (thats actually pretty good if you think about it)

The key thing is that for every 1% that I outspend my opponent, my votes increase by 0.463.

Suppose that currently I am spending the same amount as my opponent, then

$$shareA = 50$$

If I double my spending (and my opponent does not react) then

$$shareA = 67$$

I would get

$$17 \times 0.463 = 7.8$$

more of the vote

Is it worth it? (I don't know the answer to this, but this is the economic question)

CEO Example

For the CEO model we got

$$salary = 963.191 + 18.501roe + \widehat{u}_i$$

I would interpret this as the CEO's pay schedule

If the CEO can get the return on Equity to increase by 10 percentage points more, she will earn an extra $185,010 next year

This really is not that much money given how large a 10 percentage point increase is

Schooling Example

Lets think about the schooling example

In some ways income is better, but Wooldridge used hourly wage instead so lets use that

The sample regression model is

$$wage_i = \widehat{\beta}_0 + \widehat{\beta}_1 educ_i + \hat{u}_i$$

We use the data set wage1 which gives data from 1976

we got

$$wage_i = -0.90 + 0.54 educ_i + \hat{u}_i$$

For every extra year of education I get, my wage increase by about 50 cents an hour

Thus four years of education is worth

$$4 \times 0.54 = 2.16$$

an hour.

To get this into annual income figure if I am full time, I work about 2000 hours per year (50 weeks $\times$40hours per week).

This works out to about $4,320 a year

Was it worth it?

Outline

Lets completely shift gears

Imagine the following

- You have a bunch of data on x_i and y_i now
- You know the value of x^* tomorrow and want to predict what the value of y will be tomorrow

Examples

- Inflation Rate this year, Unemployment Rate next year
- Corporate Profits Today, Stock Price Tomorrow
- SAT Scores, College GPA

We are going to use the sample regression model.

In order to do prediction we need a model.

Lets once again use the linear model

$$y_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i + \hat{u}_i$$

except now we also want to define a predicted value as

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$$

Our goal is to predict

$$\widehat{y}^* = \widehat{\beta}_0 + \widehat{\beta}_1 x^*$$

It is important to note that this is a very different question than the causal model question in that WE DO NOT CHOOSE x^*

Think about the inflation/unemployment example to see the difference between the following two questions:

- I am Ben Bernanke. If I change the inflation rate to 2.3% what will happen to the unemployment rate next year
- I observe the current inflation rate is 2.3%, given that what is my best guess of the unemployment rate next year

The second question is much less ambitious (which doesn't mean it isn't hard) This second question is the forecasting question

How might I do this?

Well, I want y to be close to $\widehat{y}^*$ so for a good model we want the $\widehat{y}_i$'s "close" to the y_i's.

How do we decide what close means?

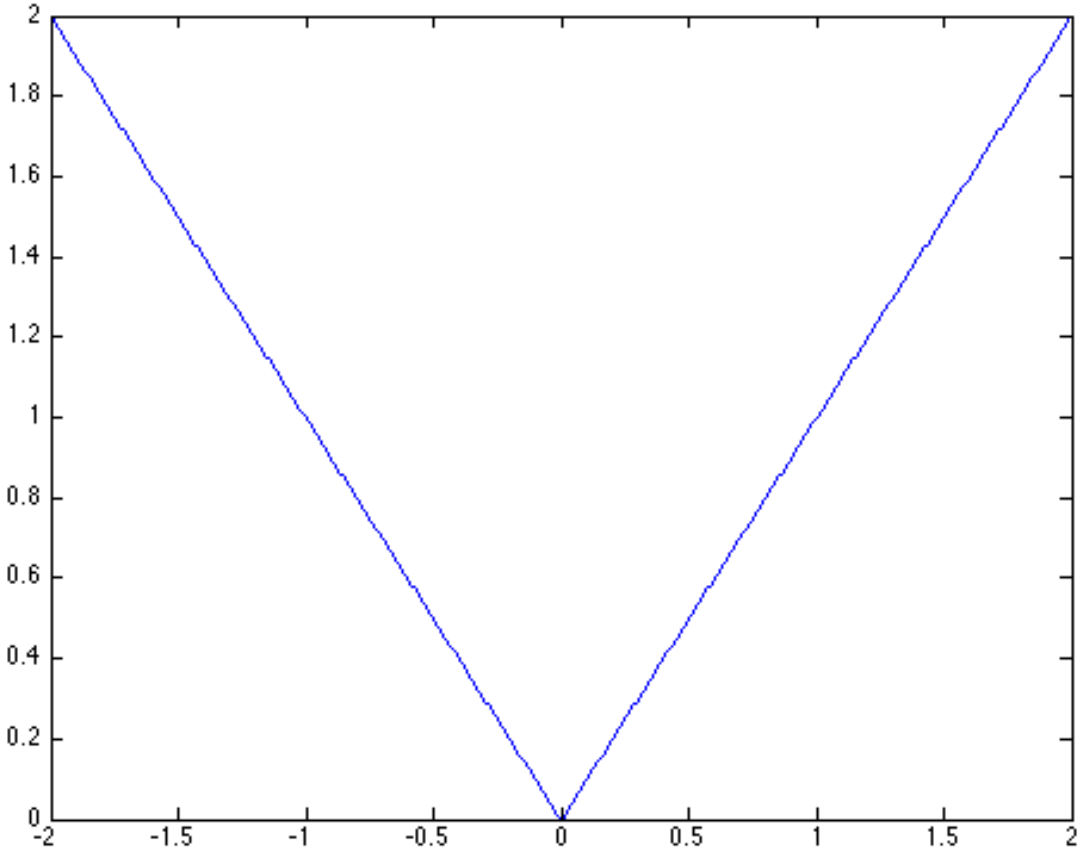
We want some function to measure the "distance" between $\hat{y}_i$ and y_i

One obvious example is just how far apart they are from each other

The first thing to try is just the absolute difference between the two

$$\left|y_i - \widehat{y}_i\right|$$

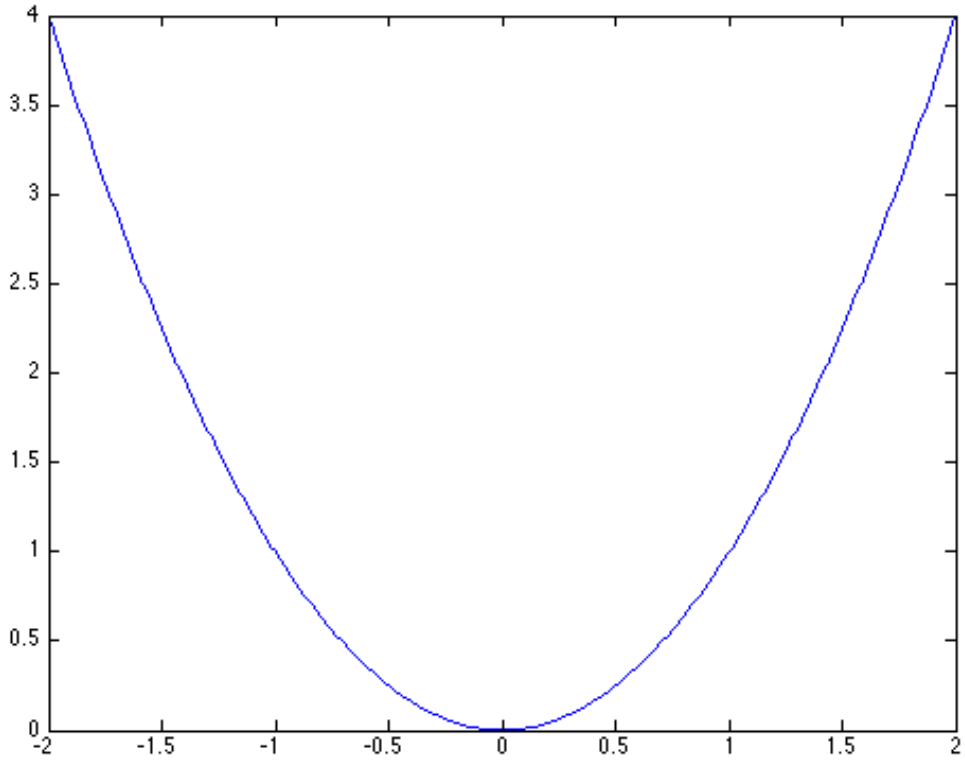
The problem is the absolute value is a really ugly function



A much smoother function is

$$(y_i - \hat{y}_i)^2$$

This looks much nicer



We have shown what the difference is for one data point, but we want to do it for all of the data points

$$\sum_{i=1}^{N} \left(y_i - \widehat{y}_i\right)^2 = \sum_{i=1}^{N} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2$$

This says how close our model is to the data.

We want it as close as possible so we want to choose $\widehat{\beta}_0$ and $\widehat{\beta}_1$ to minimize this function.

To do this we take the derivative of this function with respect to $\widehat{\beta}_0$ and $\widehat{\beta}_1$ and set to zero

First consider $\widehat{\beta}_0$,

$$
\begin{aligned}
0 &= \frac{\partial \sum_{i=1}^{N} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2}{\partial \widehat{\beta}_0} \\
&= \sum_{i=1}^{N} \frac{\partial \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2}{\partial \widehat{\beta}_0} \\
&= \sum_{i=1}^{N} -2\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) \\
&= \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) \\
&= \frac{1}{N} \sum_{i=1}^{N} y_i - \frac{1}{N} \sum_{i=1}^{N} \widehat{\beta}_0 - \frac{1}{N} \sum_{i=1}^{N} \widehat{\beta}_1 x_i \\
&= \overline{y} - \widehat{\beta}_0 - \widehat{\beta}_1 \overline{x}
\end{aligned}
$$

or

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

which is exactly what we got before

Next consider $\hat{\beta}_1$

$$
\begin{aligned}
0 &= \frac{\partial \sum_{i=1}^{N} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2}{\partial \widehat{\beta}_1} \\
&= \sum_{i=1}^{N} \frac{\partial \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2}{\partial \widehat{\beta}_1} \\
&= \sum_{i=1}^{N} -2 x_i \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} x_i \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \widehat{\beta}_0 - \frac{1}{N}\sum_{i=1}^{N} x_i \widehat{\beta}_1 x_i \\
&= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \widehat{\beta}_0 \frac{1}{N}\sum_{i=1}^{N} x_i - \widehat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i^2
\end{aligned}
$$

Now lets plug in the value for $\widehat{\beta}_0$

$$\begin{aligned} 0 &= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \left(\overline{y} - \widehat{\beta}_1 \overline{x}\right) \frac{1}{N}\sum_{i=1}^{N} x_i - \widehat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} x_i \left(y_i - \overline{y}\right) - \widehat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i \left(x_i - \overline{x}\right) \end{aligned}$$

for the same reasons as before we can write this as

$$\widehat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N} \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\frac{1}{N}\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2}$$

again exactly the same as before

Again we have another way of thinking about the same regression model.

Lets use stata to think about a few examples

- Inflation and Unemployment
- SAT scores and college GPA

Since we got

$$\widehat{\beta}_0 = 1.2792$$
$$\widehat{\beta}_1 = 0.0008918$$

Thus suppose I have two applicants, one with SAT scores of 1000, the other with 12000

what is my forecast of their GPAs

$$1.2792 + 0.0008918 \times 1000 = 2.17$$
$$1.2792 + 0.0008918 \times 1200 = 2.35$$